

Formulario de aprobación de curso de posgrado/educación permanente

Asignatura: Aprendizaje Automático para Datos en Grafos

(Si el nombre contiene siglas deberán ser aclaradas)

Modalidad:

(posgrado, educación permanente o ambas)

Posgrado

Educación permanente

Profesor de la asignatura ¹: Dr. Gonzalo Mateos (Associate Professor, University of Rochester)

(título, nombre, grado o cargo, instituto o institución)

Profesor Responsable Local ¹: Dr. Federico La Rocca (Gr. 4, IIE)

(título, nombre, grado, instituto)

Otros docentes de la Facultad: Dr. Marcelo Fiori (Gr. 3, IMERL)

(título, nombre, grado, instituto)

Docentes fuera de Facultad: No

(título, nombre, cargo, institución, país)

¹ Agregar CV si el curso se dicta por primera vez.

(Si el profesor de la asignatura no es docente de la Facultad se deberá designar un responsable local)

[Si es curso de posgrado]

Programa(s) de posgrado: Maestría y Doctorado en Ingeniería Eléctrica, Maestría en Ingeniería Matemática.

Instituto o unidad: IIE/IMERL

Departamento o área:

Horas Presenciales: 20

(se deberán discriminar las horas en el ítem Metodología de enseñanza)

Nº de Créditos: 5

[Exclusivamente para curso de posgrado]

(de acuerdo a la definición de la UdelaR, un crédito equivale a 15 horas de dedicación del estudiante según se detalla en el ítem Metodología de enseñanza)

Público objetivo: Estudiantes de programas de posgrado de Ingeniería como Maestría y Doctorado en Ingeniería Eléctrica, Maestría en Ingeniería Matemática, Maestría en Ciencia de Datos y Aprendizaje Automático o Maestría y Doctorado en Computación. Profesionales en el área de aprendizaje automático y ciencia de datos.

Cupos: No tiene.

(si corresponde, se indicará el número de plazas, mínimo y máximo y los criterios de selección. Asimismo, se adjuntará en nota aparte los fundamentos de los cupos propuestos. Si no existe indicación particular para el cupo máximo, el criterio general será el orden de inscripción, hasta completar el cupo asignado)

Objetivos:

Los grafos (o redes) son una estructura de datos presente en muchísimas áreas de conocimiento: redes de telecomunicaciones, sistemas de recomendación, redes de regulación genética, estructuras de

proteínas o movilidad urbana son solo algunos ejemplos. Básicamente, son entidades (nodos) que interactúan entre sí (aristas).

Sobre estos datos existen muchos problemas interesantes de aprendizaje automático, donde básicamente se busca realizar predicciones o descubrir cierta estructura en los datos: detección de anomalías en una red inalámbrica, recomendación de libros a partir de clasificaciones anteriores y de otros usuarios, o clasificación del rol de ciertas proteínas en redes de interacción biológica. Sin embargo, está claro que técnicas de aprendizaje "tradicionales" (donde los datos son básicamente un vector o una matriz) que no toman en cuenta las relaciones entre las distintas entidades tendrán menor poder de generalización (y por lo tanto mucho peor desempeño) que aquellas que sí lo tomen en cuenta.

El objetivo general del curso es que los estudiantes puedan afrontar un problema de aprendizaje automático donde los datos se encuentran en forma de grafos. Se brindarán los conceptos teóricos fundamentales y las herramientas prácticas necesarias para ello. Al finalizar el curso los estudiantes serán capaces de implementar y entender distintas técnicas del estado del arte en inferencia y predicción en grafos.

Conocimientos previos exigidos: Álgebra lineal y probabilidad. Programación a nivel básico.

Conocimientos previos recomendados: Conocimientos básicos de aprendizaje automático en general. Programación en python científico.

Metodología de enseñanza:

(comprende una descripción de la metodología de enseñanza y de las horas dedicadas por el estudiante a la asignatura, distribuidas en horas presenciales -de clase práctica, teórico, laboratorio, consulta, etc.- y no presenciales de trabajo personal del estudiante)

El curso se separa en cinco módulos, a los que se le dedicarán aproximadamente tres horas de clase teórica a cada uno. Se comenzará por una introducción que incluye temas fundamentales del procesamiento de señales en grafos. Luego se presentarán distintas técnicas de inferencia topológica: cómo estimar el grafo subyacente a los datos a partir de observaciones en los nodos. En tercer lugar se discutirá cómo aplicar técnicas de aprendizaje profundo en grafos, enfocándose en las denominadas Graph Neural Networks (GNNs) y sus variantes. El cuarto tema será cómo codificar el grafo (o un subconjunto) a un vector de dimensión mucho menor que el número de nodos. Finalmente, se presentarán algunas técnicas para generación de grafos. En todos los casos se buscará reflejar las técnicas con aplicaciones concretas como las ya mencionadas.

Además de la sesiones teóricas, se incluirá también un taller práctico en python que busca ilustrar algunos de los conceptos vistos en clase. Éstos serán realizados por los estudiantes y servirán sobre todo para auto-estudio y preparación del trabajo final. El objetivo es que luego de realizar el taller el estudiante sea capaz de levantar un conjunto de datos y aplicarle algunas de las técnicas que se verán en el curso, independientemente del trabajo final que elija.

Detalle de horas:

- Horas de clase (teórico): 15
- Horas de clase (práctico): 0
- Horas de clase (laboratorio): 0
- Horas de consulta: 5
- Horas de evaluación: 0
 - Subtotal de horas presenciales: 20

- Horas de estudio: 10
- Horas de resolución de ejercicios/prácticos: 10
- Horas proyecto final/monografía: 35
 - Total de horas de dedicación del estudiante: 75

Forma de evaluación:

Un proyecto final en el que los estudiantes podrán investigar y aplicar algoritmos de aprendizaje automático en grafos del estado del arte, a una aplicación de su interés.

Una vez concluido el componente instructivo del curso, los estudiantes deben seleccionar un tema específico, relacionado con el material tratado en clase, y realizar un estudio relativamente profundo del tema. Esto implica por ejemplo buscar fuentes de literatura (trabajos de investigación, libros de texto, etc.), y eventualmente realizar algunos análisis y/o simulaciones numéricas para experimentar con conjuntos de datos sobre grafos, y proporcionar un resumen detallado de las ideas principales en formato paper.

Temario:

- **Introducción**, motivación, graph signal processing (GSP). Introducción al aprendizaje automático en grafos. Aplicaciones motivadoras y desafíos. Preliminares teóricos de grafos. Conceptos básicos de GSP: graph-shift operator; graph Fourier transform (GFT); y filtros/convoluciones en grafos.
- **Graph learning**. El problema de inferencia de topología. Métodos estadísticos (grafos de correlación, selección de covarianza con graphical lasso, y regresión basada en vecinos). Aprendizaje de grafos a partir de observaciones de señales: análisis basados en Laplaciano, suavidad y esparsidad. Identificación de estructura en procesos de difusión. Aplicación a redes de regulación genética, estructura de proteínas, y patrones de movilidad urbana.
- **Graph neural networks (GNNs)**. Arquitecturas de GNN como composición de filtros en grafos con no-linealidades. Grafos en el dominio del espectro; ChebNet; CayleyNet; redes convolucionales en grafos. Propiedades fundamentales de las GNN: invariancia a permutaciones, estabilidad a deformaciones de los grafos, expresividad y poder de discriminación. Variantes: graph attention networks; recurrent GNNs; pooling. Aplicaciones a predicción de aristas, aprendizaje semi-supervisado y sistemas de recomendación.
- **Graph representation learning**. Detección de comunidades, Laplacian embeddings y spectral clustering. Node y (sub)graph representation learning: métodos basados en factorización de matrices, paseos aleatorios, y redes convolucionales en grafos. Aplicaciones a clasificación de nodos y de grafos, y modelado de la relación entre conectividad estructural y funcional en el cerebro.
- **Modelos generativos**. Modelos clásicos, Erdos-Renyi, aplicaciones a estimación basada en modelos, detección de motifs, crecimiento de grafos por preferential attachment. Generación de grafos a través de Deep generative models: graph variational autoencoders; generative adversarial networks.

Bibliografía:

- Statistical Analysis of Network Data Methods and Models, Eric Kolaczyk, Springer, 2009.
 - Deep Learning, Ian Goodfellow and Yoshua Bengio and Aaron Courville, MIT Press, 2016.
- Disponible en línea: <https://www.deeplearningbook.org/>.

- Ortega, P. Frossard, J. Kovačević, J. M. F. Moura and P. Vandergheynst, "Graph Signal Processing: Overview, Challenges, and Applications," in Proceedings of the IEEE, vol. 106, no. 5, pp. 808-828, May 2018.
- G. Mateos, S. Segarra, A. G. Marques and A. Ribeiro, "Connecting the Dots: Identifying Network Structure via Graph Signal Processing," in IEEE Signal Processing Magazine, vol. 36, no. 3, pp. 16-43, May 2019.
- X. Dong, D. Thanou, M. Rabbat and P. Frossard, "Learning Graphs From Data: A Signal Representation Perspective," in IEEE Signal Processing Magazine, vol. 36, no. 3, pp. 44-63, May 2019.
- M. M. Bronstein, J. Bruna, Y. LeCun, A. Szlam and P. Vandergheynst, "Geometric Deep Learning: Going beyond Euclidean data," in IEEE Signal Processing Magazine, vol. 34, no. 4, pp. 18-42, July 2017.
- J. Zhou et al, "Graph Neural Networks: A Review of Methods and Applications", arXiv:1812.08434v4 [cs.LG].
- F. Gama, E. Isufi, G. Leus, A. Ribeiro, "Graphs, Convolutions, and Neural Networks," arXiv:2003.03777v1 [cs.LG].
- T. N. Kipf, M. Welling, "Semi-Supervised Classification with Graph Convolutional Networks," arXiv:1609.02907v4 [cs.LG].
- W. L. Hamilton, R. Ying, J. Leskovec, "Representation Learning on Graphs: Methods and Applications," arXiv:1709.05584v3 [cs.SI].
- T. N. Kipf, M. Welling, Variational Graph Auto-Encoders, NIPS Workshop on Bayesian Deep Learning (2016).
- J. You, R. Ying, X. Ren, W. L. Hamilton, J. Leskovec, "GraphRNN: Generating Realistic Graphs with Deep Auto-regressive Model", ICML 2018.
- William L. Hamilton, "Graph Representation Learning", Morgan & Claypool, 2020. Borrador disponible en línea: https://www.cs.mcgill.ca/~wlh/grl_book/.

(título del libro-nombre del autor-editorial-ISBN-fecha de edición)

Datos del curso

Fecha de inicio y finalización: El curso inicia el lunes 1º de febrero de 2021 y finaliza el viernes 5 de febrero.

Horario y Salón: a confirmar.

Arancel:

[Si la modalidad no corresponde indique "no corresponde". Si el curso contempla otorgar becas, indíquelo]

Arancel para estudiantes inscriptos en la modalidad posgrado: No corresponde

Arancel para estudiantes inscriptos en la modalidad educación permanente: 2.500 UI
